



Functional Prediction and Assignment of *Methanobrevibacter ruminantium* M1 Operome Using a Combined Bioinformatics Approach

M. Bharathi¹, N. Senthil Kumar² and P. Chellapandi^{1*}

¹ Molecular Systems Engineering Lab, Department of Bioinformatics, School of Life Sciences, Bharathidasan University, Tiruchirappalli, India, ² Human Genetics Lab, Department of Biotechnology, School of Life Sciences, Mizoram University (Central University), Aizawl, India

OPEN ACCESS

Edited by:

Saumya Patel,
Gujarat University, India

Reviewed by:

Khanh N. Q. Le,
Taipei Medical University, Taiwan
Sailu Yellaboina,
CR Rao Advanced Institute
of Mathematics, Statistics
and Computer Science, India

*Correspondence:

P. Chellapandi
pchellapandi@gmail.com

Specialty section:

This article was submitted to
Computational Genomics,
a section of the journal
Frontiers in Genetics

Received: 12 August 2020

Accepted: 17 November 2020

Published: 16 December 2020

Citation:

Bharathi M, Senthil Kumar N and
Chellapandi P (2020) Functional
Prediction and Assignment
of *Methanobrevibacter ruminantium*
M1 Operome Using a Combined
Bioinformatics Approach.
Front. Genet. 11:593990.
doi: 10.3389/fgene.2020.593990

Methanobrevibacter ruminantium M1 (MRU) is a rod-shaped rumen methanogen with the ability to use H₂ and CO₂, and formate as substrates for methane formation in the ruminants. Enteric methane emitted from this organism can also be influential to the loss of dietary energy in ruminants and humans. To date, there is no successful technology to reduce methane due to a lack of knowledge on its molecular machinery and 73% conserved hypothetical proteins (HPs; operome) whose functions are still not ascertained perceptively. To address this issue, we have predicted and assigned a precise function to HPs and categorize them as metabolic enzymes, binding proteins, and transport proteins using a combined bioinformatics approach. The results of our study show that 257 (34%) HPs have well-defined functions and contributed essential roles in its growth physiology and host adaptation. The genome-neighborhood analysis identified 6 operon-like clusters such as *hsp*, TRAM, *dsr*, *cbs* and *cas*, which are responsible for protein folding, sudden heat-shock, host defense, and protection against the toxicities in the rumen. The functions predicted from MRU operome comprised of 96 metabolic enzymes with 17 metabolic subsystems, 31 transcriptional regulators, 23 transport, and 11 binding proteins. Functional annotation of its operome is thus more imperative to unravel the molecular and cellular machinery at the systems-level. The functional assignment of its operome would advance strategies to develop new anti-methanogenic targets to mitigate methane production. Hence, our approach provides new insight into the understanding of its growth physiology and lifestyle in the ruminants and also to reduce anthropogenic greenhouse gas emissions worldwide.

Keywords: methanobrevibacter, methane mitigation, hypothetical proteins, protein function, molecular machinery

INTRODUCTION

Enteric methane emission from ruminants is of great concern not only for its impact on global warming potential but also for ensuring the long-term sustainability of ruminant-based agriculture. Methane emission from rumen methanogens (163.3 million metric tons of CO₂ equivalents) represents a loss of about 5–7% of dietary energy in ruminants (Hristov et al., 2013;

Chellapandi et al., 2017a, 2018; Chellapandi and Prathiviraj, 2020). *Methanobrevibacter* genus is a dominant rumen methanogenic archaea (61.6%) in which *Methanobrevibacter ruminantium* M1 (MRU) accounted for 27.3% (Janssen and Kirs, 2008). MRU is a hydrogenotrophic rumen methanogen that use H₂ to reduce CO₂ for methane biosynthesis. It also uses formate as a carbon source for its growth and energy metabolism (Kaster et al., 2011). This is the first genome sequence to be completed for rumen methanogen. It is a circular chromosome (2.93 Mbp) consisting of 2,278 coding-genes and 144 metabolic pathways with 722 reactions, 557 enzymes, and 751 metabolites (Leahy et al., 2010). However, the MRU genome consists of 756 coding-genes (73%) annotated as hypothetical proteins (HPs). It suggests that the entire proteome functions of this organism are not yet known and have to be elucidated to date.

The function of only 50–70% of coding-genes has been annotated with reasonable confidence in the most completely sequenced bacterial genomes using automated genome sequence analysis (Loewenstein et al., 2009). The characterization of proteins with unknown biological function is known as operome (Greenbaum et al., 2001; Chellapandi et al., 2017b; Prathiviraj and Chellapandi, 2019). Putative genes with known orthologs and no orthologs are termed as conserved hypothetical proteins and uncharacterized proteins, respectively (Mazandu and Mulder, 2012; Shahbaaz et al., 2013). Several approaches have been developed for assisting the function of operome from prokaryotic genomes using the information derived from sequence and structural motifs (Sivashankari and Shanmughavel, 2006; Chellapandi et al., 2017b; Singh and Singh, 2018; Prathiviraj and Chellapandi, 2020a; Sangavai et al., 2020). No one has been employed a combined bioinformatics prediction approach including sequence, structure, and literature confidences for functional assignment of operome and its contribution to metabolic subsystems and cellular machinery. A precise annotation of the operome of a particular genome leads to the discovery of new functions for the development of veterinary and human therapeutics (Ijaq et al., 2015).

The conserved domain-based functional assignment was done for HPs from *Pongo abelii* and *Sus scrofa*. It has provided a hint for genome-wide annotation in poorly understood genomes (Jitendra et al., 2011). The structure-based approach has been applied to predict the function of operome from *Mycoplasma hyopneumoniae* (da Fonsêca et al., 2012). Functional and structural domain analysis (Namboori et al., 2004), integrated genomic context analysis (Yellaboina et al., 2007) and literature mining (Doerks et al., 2012), functional enrichment analysis (Mazandu and Mulder, 2012), and genome-scale fold-recognition (Mao et al., 2013) have been used to annotate the potential function of operome from *Mycobacterium tuberculosis* H37Rv. Sequence-based and structure-based approaches have been used to define and prioritize some HPs from *Candida dubliniensis*, *Vibrio cholerae* O139, and *Staphylococcus aureus* as therapeutic targets for the treatment of their infections in humans (McAdow et al., 2011, 2012; Bharat Siva Varma et al., 2015; Islam et al., 2015). Besides, only one HP (MJ_0577) was functionally annotated in *Methanococcus jannaschii* using a structural-based approach (Zarembinski et al., 1998).

Many *in silico* attempts have been focused on the functional prediction of operome from human pathogens and no reports on rumen methanogens. Several genome-scale metabolic networks have been reconstructed for methanogenic archaea with a low fraction of HPs functionally assigned by sequence similarity analysis (Chellapandi et al., 2018; Prathiviraj and Chellapandi, 2020a). Since, functional annotation of operome is a great concern not only for implementing our fragmentary knowledge on the potential drug targets but also for genome refinement and improved microbial genome-scale reconstructions (Poulsen et al., 2010; Mazandu and Mulder, 2012; Prathiviraj and Chellapandi, 2019). Thus, we have employed a combined bioinformatics approach for functional assignment, and categorization of operome from MRU with a biological knowledgebase. The predicted functions of operome allow us to comprehend its growth physiology and metabolic behavior in the rumen environment. Several methanogenic antibiotics, inhibitors, and vaccines have been currently available for enteric methane mitigation, but these are a narrow spectrum and species-specific activity (Pulendran and Ahmed, 2006). The present approach is used to predict new anti-methanogenic targets from its precisely annotated operome that resolves the current demand for veterinary therapeutics.

MATERIALS AND METHODS

Dataset Preparation

We retrieved protein sequences of 756 HPs in the MRU genome from the National Centre for Biotechnology Information (NCBI)¹ and Kyoto Encyclopedia of Genes and Genomes (KEGG) (Kanehisa et al., 2018) using a simple text mining approach (Le and Huynh, 2019; Le et al., 2019). We used broad ranges of source types such as keywords, “*hypothetical proteins, unknown, uncharacterized, and putative*” to retrieve the protein sequences from the NCBI and KEGG (Chellapandi et al., 2017b). The FASTA sequences of all HPs were taken separately to carry out sequence analysis. For functional annotation and assignment of MRU operome, we used six different prediction tasks as detailed below (Figure 1). The overall information about similar or identical functions of HPs predicted from each task was manually evaluated to reasoning out the functional assignment of operome. The prediction tools used for each functional annotation were more robust and confident for our analysis similar to the previous works on archaeal and bacterial operome (Prathiviraj and Chellapandi, 2019; Sangavai et al., 2020). E-value is the number of expected hits of a similar score that could be found just by chance. Like *p*-value, we used e-value for the scoring of each prediction from the dataset and represented in **Supplementary Data**.

Conserved Motif Analysis

A motif is a short segment of a protein sequence or structure, which may be conserved in a large number of different proteins. It can be used to determine the function or conformation

¹<http://www.ncbi.nlm.nih.gov>

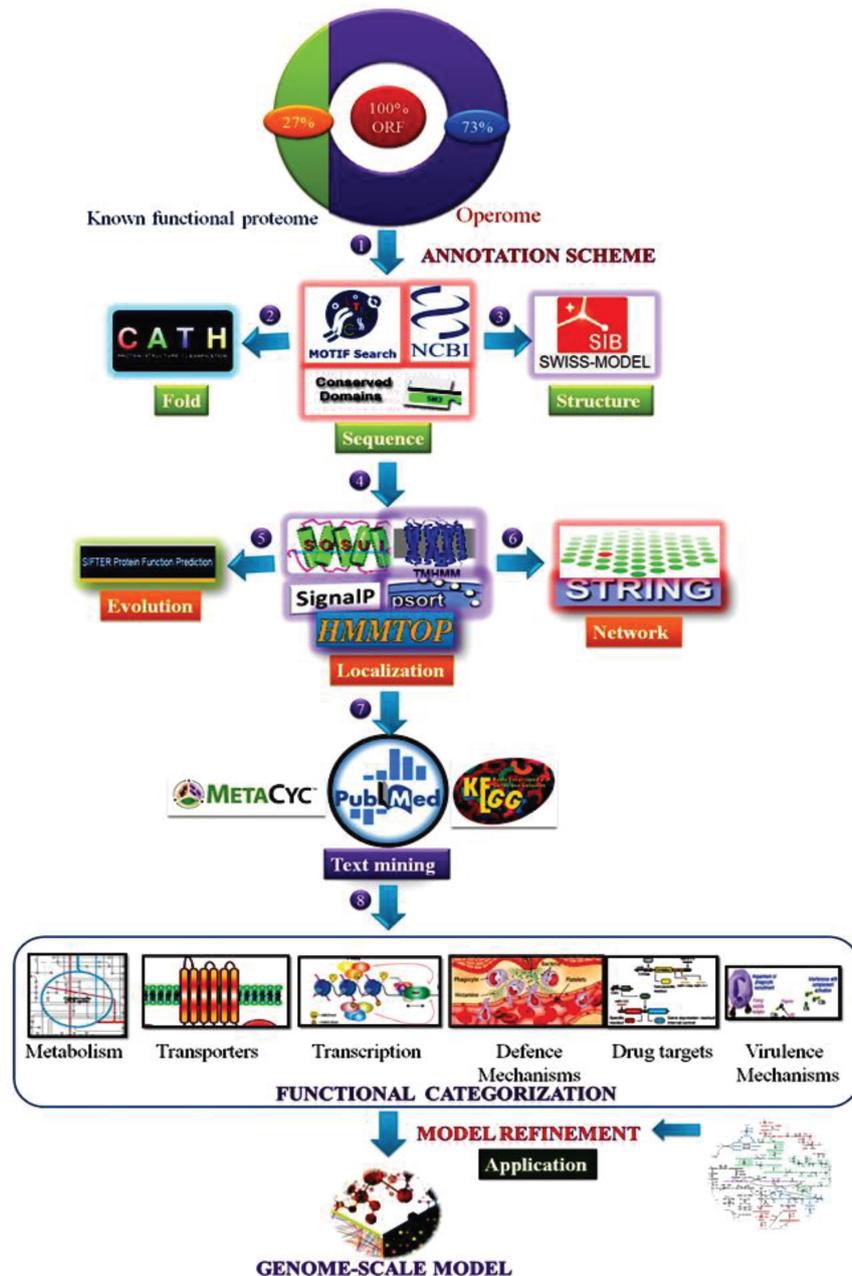


FIGURE 1 | Experimental workflow of a combined bioinformatics approach employed for functional annotation of operome from MRU.

of a protein. The conserved motifs in each protein were searched out against the KEGG-Motif search tool², InterProScan (Quevillon et al., 2005), and Pfam library (Finn et al., 2016). To improve the lineament of prediction, cut off value was set as 10^{-5} and DUF (domains with unknown functions) were removed from the dataset. We found motif similarity hits for 756 HPs out of which 257 HPs were chosen for further analysis.

²<http://www.genome.jp/tools/motif/>

Conserved Domain Analysis

Conserved domains in each protein were identified by the NCBI-CDD v3.16 search tool using the position-dependent weight matrices. Additionally, composition-based statistics adjustment was used to remove low complexity composition for statistical significance using the RPS-BLAST version 2.2.28 (Marchler-Bauer et al., 2015). The query sequence was compared with domain architecture and profiles in the domain databases, after that, the compositionally biased conserved region was identified by the SMART (Letunic et al., 2012). The PROSITE profile

was scanned for detection of the protein domains, families, and functional sites and associated patterns in the protein sequence using ScanProsite (de Castro et al., 2006). The probable function of HPs was predicted with the InterPro database based on the domain and important sites in the sequences (Finn et al., 2016).

Structural Analysis

The secondary structural elements (helix, sheets, extended coil, and loops) in each protein were predicted from the sequences using SOPMA (Geourjon and Deléage, 1995). We identified structural and functional characteristics by PSI-BLAST similarity searching against the protein data bank³ (Altschul et al., 1997). The sequence similarity hits were selected for finding the alignment of functional residues of a protein of known function with the sequence of HPs using ClustalW (Thompson et al., 2002). Fold assignment, target-template alignment, model building, and model evaluation were carried out with the Swiss Model (Biasini et al., 2014). QMEAN was a composite scoring function describing the major geometrical aspects of protein structures as described below.

$$S_{\text{weighted average}(x)} = \frac{\sum_i (\text{GDT}_{\text{TS}(x,i)}^* \text{QMEAN}(i))}{\sum_i \text{QMEAN}(i)}$$

where, the GDT_TS score as the target function. We evaluated the structural quality and accuracy of the resulted homology models based on the potential function as below (Benkert et al., 2008).

$$\begin{aligned} \text{QMEAN5 score} &= 0.3 \times \text{Score}_{\text{torsion } 3\text{-residue}} + 0.17 \times \\ &\text{Score}_{\text{pairwise } C\beta/\text{SSE}} + 0.7 \times \\ &\text{Score}_{\text{solvation } C\beta} + 80 \times \text{Score}_{\text{SSE PSIRED}} \\ &+ 45 \times \text{Score}_{\text{ACCpro}} \end{aligned}$$

Evolutionary Trace Analysis

The evolutionary relationships to deduce the functionality of operome were inferred using the SIFTER (Radivojac et al., 2013). It was used to predict the protein function and Gene ontology term using the following confidence score.

$$\text{Sg}(f) = 1 - \prod_{i=1}^k (1 - \text{Sg}_i(f))$$

where, Sg(f) confidence score as the default prediction for a query protein g, Sg_i(f) is the probability domain has function f (Sahraeian et al., 2015).

Analysis of Physicochemical Properties

The physicochemical properties including molecular weight, theoretical pI, instability index, aliphatic index, and grand average of hydropathicity of HPs were predicted from their sequences using the Expasy's ProtParam server⁴. The instability index provides an estimate of the stability of a protein. An

instability index <40 is predicted to be stable, and a value >40 is predicted to be unstable. The instability index uses the following weight values.

$$\Pi = \left(\frac{10}{L}\right) * \sum_{i=1}^{i=L-1} \text{DIWV}(x[i]x[i+1])$$

where, L is the length of the sequence, DIWV(x[i]x[i+1]) is the instability weight value for the dipeptide starting in position I (Guruprasad et al., 1990). The aliphatic index of a protein is defined as the relative volume occupied by aliphatic side chain amino acids using the following equation.

$$\text{Aliphatic index} = X(\text{Ala}) + a * X(\text{Val}) + b * (X(\text{Ile}) + X(\text{Leu}))$$

Where, X(Ala), X(Val), X(Ile), and X(Leu) are mole percent (100 X mole fraction) (Ikai, 1980). The GRAVY value for a protein is calculated as the sum of the hydropathy values of all of the amino acids divided by the number of residues in the sequence (Kyte and Doolittle, 1982).

Analysis of Protein Subcellular Localization

The subcellular localization of every protein was predicted with PSORTb version 3.0.2 based on the hydrophobicity index of amino acids (Yu et al., 2010). The propensity of a protein for being a membrane protein was predicted by SOSUI 2.0 based on the physicochemical parameters (Mitaku et al., 2002). The transmembrane helix and topology of each protein were detected by the TMHMM 2.0 (Krogh et al., 2001) and HMMTOP (Tusnády and Simon, 2001) using the Hidden Markov Model. The signal peptide and location of the cleavage site in the peptide chain were predicted with the SignalP 4.0 based on a neural network model (Petersen et al., 2011).

Literature Search

The literature survey is the stepping-stone and an essential skill toward the accomplishment of structural and functional analysis provides of proteins (Hubbard and Dunbar, 2017). A process of uncovering useful knowledge from a collection of data from bioinformatics and literature databases is referred to as a knowledge-based discovery (Chellapandi et al., 2017b). Functional assessment of operome was strengthened by extracting relevant experimental supports from available literature in NCBI-PubMed⁵. A maximum confidence score was set as 12 levels (6 levels from predictions and 6 levels from the literature mining) in which 50% score systematically enumerated and assigned from overall prediction approaches. The rest of them was assigned by manual annotation based on the strength of the literature validation. For example, if the predicted function is similar or identical in all prediction approaches, a maximum confidence score will be assigned as 6. The literature-based confidence score for each predicted function of HPs assigned as; 6- MRU, 5- Phylogenetic neighbors, 4- Methanogens, 3- Archaea,

³www.rcsb.org/

⁴http://web.expasy.org/protparam/

⁵https://www.ncbi.nlm.nih.gov/pubmed/

2- Bacteria, and 1- Eukaryotes. We have set a confidence score interval as 3–6 for both computational prediction and biological knowledge base and then neglected the predicted function of a protein with a low confidence score (<3).

Functional Categorization

We classified the predicted function of HPs based on conserved domain, protein fold, family, and biological function using the CATH database (Knudsen and Wiuf, 2010). The genome-wide analysis was performed to identify the order of gene clusters covering the predicted function of HPs using a genomic context approach (Yellaboina et al., 2007). Gene-neighborhood or adjutant genes were identified by exploring the MRU genome in the KEGG database. Metabolic information of HPs was collected from the MetaCyc (Metabolic Pathways from all Domains of Life) database (Caspi et al., 2014). The resulted data were used to assign the functions of hypothetical proteins of the understudied genome. The overall structural and functional information was manually analyzed to categorize the molecular involvement of HPs in respective metabolic subsystems and the cellular process of the understudied organism.

RESULTS

Functional Classification and Categorization

All predicted protein functions were classified and categorized according to their protein folds, molecular function, subsystems, and transmembrane topologies as shown in **Figure 2**. About 20% of operome encompasses a Rossmann fold consisting of a nucleotide cofactor binding domain of some NAD⁺-dependent dehydrogenases, in particular to ribonucleases (Barbas et al., 2013). Fourteen percent of operome belongs to rubrerythrin that constitutes non-haem iron proteins. This functional fold is responsible for oxidative stress protection in anaerobic bacteria and archaea (Prakash et al., 2018). The arcR repressor mutant fold occupies 4–5% of operome, which performs the functions of small homodimeric proteins involved in transcriptional regulation by sequence-specific DNA binding (Vershon et al., 1986; Homa and Brown, 1997). MRU operome contains phoA fold (3–4%) that fused with the cell surface glycoprotein signal sequence similar to *Haloferax volcanii* (Kandiba et al., 2013). It indicates the importance of some protein folds for conferring oxidative tolerance and cell wall assembly. We found 91 HPs involving in the metabolic reactions with a confidence score >5. A total of 23 HPs is entailed in the small molecule reactions and 15 HPs required for the biosynthesis of cofactors, prosthetic groups, and electron carriers. About 9 HPs are essential to the protein modification reactions whereas 4 HPs contributed to the formation of precursor metabolites for the energy-driven process of this organism. Approximately 50% of drug targets are transmembrane proteins as they play many roles in transport, cell signaling, and energy transduction processes (Terstappen and Reggiani, 2001). We predicted 91 HPs having transmembrane helices based on their conservation of membranous helix ratios. The α -helix bundle and the β -barrel are

predicted as fold classes in many membrane proteins. Archaeal transmembrane proteins have two or more α -helices consisting of hydrophobic amino acids.

Operon-Like Organization

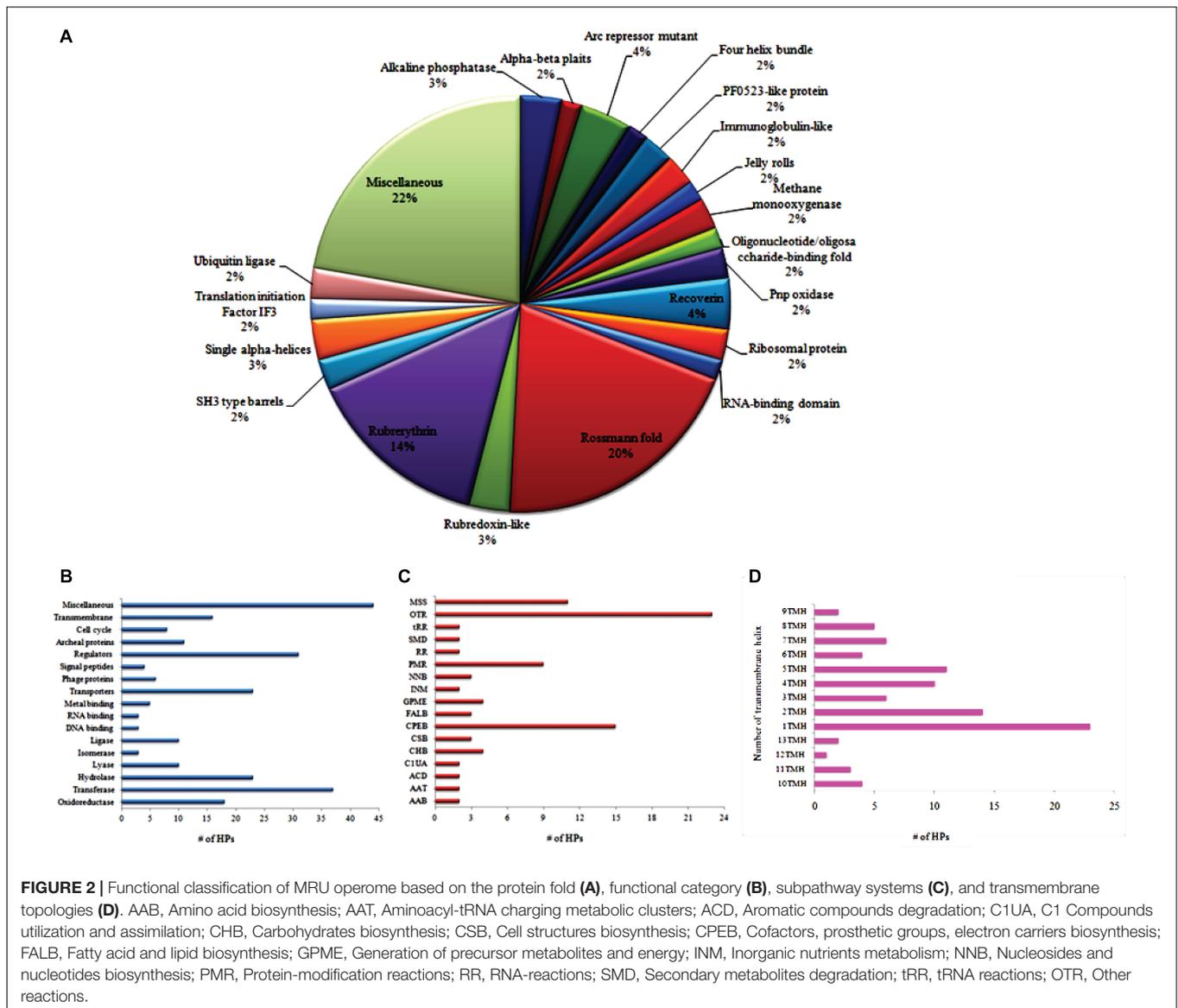
The genome-wide analysis discovered 32 coding genes for HPs, which are all clustered separately, form 6 operon-like organizations (*hsp*, *TRAM*, *dsr*, *cbs*, *anti-toxin*, and *cas*) in the MRU genome (**Figure 3**). Molecular chaperones such as hsp70, hsp60, and hsp80 resemble some bacterial genomes than the eukaryotic homologs (Gaywee et al., 2002). The *hsp* gene cluster is essential for chaperone-assisted protein folding in Achaea (Dokland, 1999; Benaroudj and Goldberg, 2000; Large et al., 2009). The assimilatory sulfite reductase (*dsrHFEBA*) gene cluster detected from this genome provides the importance for the oxidation of accumulated intracellular sulfide and thiosulfate in the diverse environmental niche. The presence of *cbs*, *anti-toxin*, and *cas* gene clusters confers host defense response (innate immunity) to this organism against foreign genetic elements in the rumen ecosystem (Louwen et al., 2014; Chellapandi and Ranjani, 2015). The anti-toxin system plays a vital role in toxicity neutralization (Unterholzner et al., 2013).

Cell Division Systems

In this study, we assigned the function of 9 HPs contributing a major role in the cell cycle process in which 8 HPs have shown new functions to this organism (**Table 1**). AAA⁺ ATPase, cell division inhibitor, cell division control protein, DNA replication protein 6-2, and structural maintenance of chromosomes protein-1 is highly conserved within the archaeal domain and performs archaeal-specific cell cycle process, DNA repair, and replication fidelity (Kallioma-Sanford et al., 2012; Grogan, 2015). A proteasome is a central player in energy-dependent proteolysis and forms a nano-compartment where proteins are degraded into oligopeptides by processive hydrolysis. The 20S proteasome is a catalytic core responsible for this processing. AAA⁺ ATPase plays several roles in mediating energy-dependent proteolysis by the proteasome (Forouhar et al., 2011; Maupin-Furlow, 2013). Moreover, it contains a P-loop motif involved in the origin of recognition during DNA replication initiation even if conventional C-terminal winged-helix DNA-binding elements lacked (He et al., 2008).

Transcriptional Regulatory Systems

A total of 26 HPs predicted as functional candidates in which 20 HPs have shown new functions to the transcriptional regulation process of this organism (**Table 2**). Transcriptional regulatory proteins identified from MRU operome can express a set of proteins that protect cellular proteins against a sudden heat-shock stress, copper and arsenic toxicities, protein folding, and nitrogen starvation (Thieringer et al., 1998; Giaquinto et al., 2007; Chang et al., 2014; Prathiviraj and Chellapandi, 2020a,b). Bro N-terminal domain protein has an N-terminal domain with ALI motif that influences host DNA replication and/or transcription (Makarova et al., 2009). HrcA repressor contains a motif of winged helix-turn-helix transcription repressor. It controls the transcription of heat-shock repressor proteins and



protects cellular proteins from being denatured by heat (Liu et al., 2005; Prathiviraj and Chellapandi, 2020b). Hsp70 and Hsp80 from MRU operome perform renaturation of luciferase similar to that found in *M. mazei* (Zmijewski et al., 2004). Hsp60s are more similar to the type II chaperonins found in the eukaryotic cytosol involved in macromolecular assembly and protein folding (Large et al., 2009). TRAM protein regulates the RNA chaperone activity that is essential for MRU to grow and survive in a cold environment (Zhang et al., 2017).

Biosynthesis of Macromolecules

We predicted the function of 20 HPs exhibiting new metabolic roles in this organism and the rest of 76 HPs has shown known functions (Table 3 and Supplementary Table S1). Saccharopine dehydrogenase (NAD/P, L-lysine-forming) (*lysA*) and succinylglutamate desuccinylase (*astE*) genes identified from MRU operome, which are responsible to mediate the

biosynthesis of L-lysine and L-glutamate. *LysA* protein contains a motif of LOR/SDH bifunctional conserved region that converts L-saccharopine into L-lysine via l- α -amino adipate pathway (Xu et al., 2007). Cheng et al. (2010), revealed a cross-talk between fungi and methanogens which may occur in host animals since the l- α -amino adipate pathway is very specific to fungi. The second enzyme transforms N₂-succinylglutamate into succinate and glutamate. Therefore, both enzymes proposed to be involved in amino acid biosynthesis of MRU as reported earlier on other methanogens (Enzmann et al., 2018).

The 2-enoyl-CoA hydratase catalyzes the second step in the physiologically important β -oxidation pathway of fatty acid metabolism in MRU (Agnihotri and Liu, 2003). Glycogen phosphorylase catalyzes the phosphorolysis of α -1, 4 glycosidic bonds in glycogen to yield glucose-1-phosphate for glycolysis (Rath et al., 2000). Interestingly, MRU operome has the ability to synthesis enterobacterial-like common

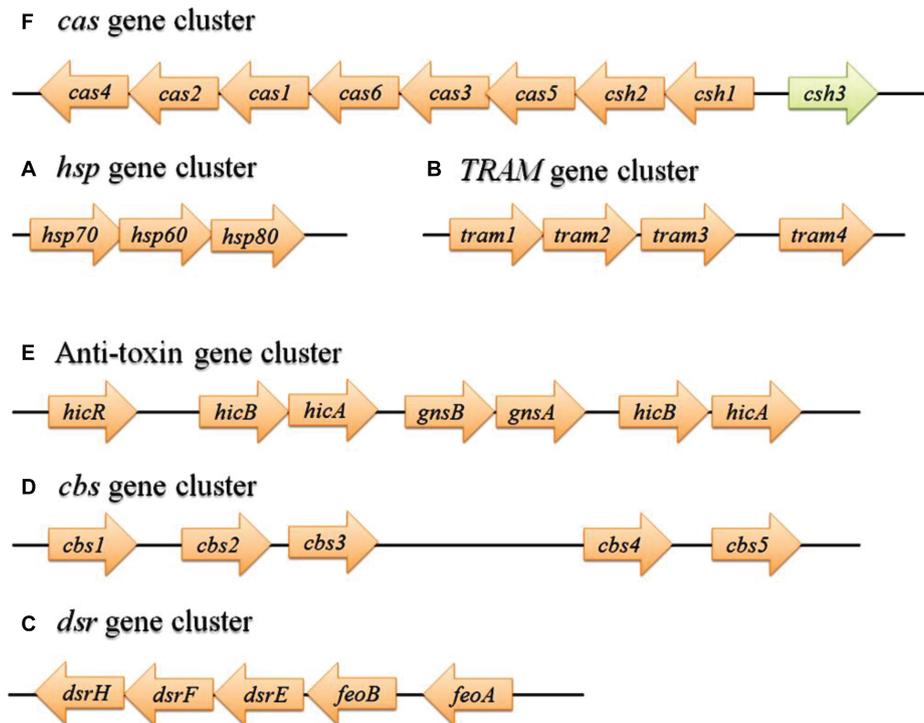


FIGURE 3 | Detection of gene clusters from MRU operome responsible for protein folding (A), cold adaptation (B), sulfite tolerance (C), binding with adenosyl groups (D), degradation of the labile antitoxin (E), and defense/virulence system (F). The green arrow represents a gene with a known function. hsp, Heat shock protein; TRAM, RNA modification protein; dsr, Dissimilatory sulfate reductase; cbs, cystathionine beta-synthase; cas, CRISPR-associated gene.

antigen as it contains dTDP-4-amino-4, 6-dideoxygalactose transaminase (*rffA*). This enzyme catalyzes the conversion of TDP-4-keto-6-deoxy-D-glucose to TDP-D-fucosamine similar to the enterobacteria family (Meier-Dieter et al., 1990; Hwang et al., 2004). The presence of phosphatidate cytidyltransferase (*cdsA*) provides evidence of the biosynthesis of archaeal-specific phospholipids. It catalyzes *sn*-glycerol 3-phosphate into an L-1-phosphatidylglycerol-phosphate precursor-like *Escherichia coli* (Carter et al., 1968). We found an AMMECR1 motif in phosphomevalonate decarboxylase from MRU operome, which converts (R)-mevalonate 5-phosphate to isopentenyl diphosphate in the mevalonate pathway, as reported in *Methanocaldococcus jannaschii* (Grochowski et al., 2006). Results of our study revealed that the MRU genome has shown a metabolic potential for the biosynthesis of enterobacterial-like common antigen, archaeal-specific phospholipids, and isopentenyl diphosphate, a precursor required for cell wall biogenesis.

Cofactors, Prosthetic Groups, Electron Carrier Biosynthesis

We predicted the function of some HPs involving in the biosynthesis of coenzyme F₄₂₀, flavin, and electron carriers in MRU. F₄₂₀-0: L-glutamate ligase is a key enzyme identified from MRU operome, which converts multiple γ -linked L-glutamates to the polyglutamated F₄₂₀ derivative in the

biosynthesis of coenzyme F₄₂₀ (Li et al., 2003). As reported in bacteria and plants, MRU operome has diamino hydroxy phosphoribosyl aminopyrimidine reductase (*ribD*) that converts 2, 5-diamino-6-(5-phospho-D-ribosylamino)pyrimidine-4(3H)-one into 5-amino-6-(5-phospho-D-ribosylamino)uracil in flavin biosynthesis pathway (Garfoot et al., 2014). Cytidyltransferase belongs to the NTP transferase superfamily encoded by *mocA* gene (*mru_1116*) of the MRU genome. It catalyzes the cytidylation of the molybdenum cofactor demanded many functional enzymes (Fay et al., 2015). Energy-converting hydrogenase B subunit O consists of a conserved motif of IHPPAH, which generates low potential electrons required for autotrophic CO₂ assimilation as reported in *Methanococcus maripaludis* (Major et al., 2010).

Aromatic Compounds Degradation Systems

Pyrogallol hydroxytransferase (*athL*) detected from MRU operome has a carboxypeptidase regulatory-like domain. It is involved only in the regulation of peptidase catalyzing the conversion of pyrogallol into phloroglucinol. Phloroglucinol can stimulate the gut microbiota and decreases the partial pressure of H₂ in the rumen. It suggests the capture of excess H₂ generated from methanogenesis inhibition can be promoted by phloroglucinol utilization in the rumen (Martinez-Fernandez et al., 2017). Interestingly, we assigned a precise function to HP

TABLE 1 | Functional annotation of operome involved in the cell division process of MRU.

Locus tag	Assigned function	Gene
0080 0744 0939 1172 1932	AAA ⁺ ATPase	<i>atad3A</i>
0647	Cell division inhibitor	<i>sepF</i>
1346	Cell division control protein	<i>minE</i>
1419	DNA replication protein 6-2	<i>cdc6-2</i>
1654	Structural maintenance of chromosomes protein 1	<i>smc1</i>

TABLE 2 | Functional annotation of operome involved in the transcriptional regulatory process of MRU.

Locus tag	Assigned function	Gene
0757	Bro N-terminal domain protein	<i>dxs</i>
0349	Nitrogen repressor	<i>nrpR</i>
1052	Heat-inducible transcriptional repressor	<i>hrcA</i>
1099	Translation initiation factor 3	<i>tif3</i>
1366 2156	Arsenical resistance operon repressor	<i>arsR</i>
1862	Copper-sensing transcriptional repressor	<i>csor</i>
0488 0490 0499 0658 0764 0780 0790 0801 0930 1131 1147 1150 1364 1590 1796	Transcription factor	<i>tf2B</i>
1185	Cold shock protein	
1108	DEAD/DEAH box helicase	<i>polB</i>
0877	Preprotein translocase	<i>secY</i>

Mru_0476 as phenylacetate-CoA oxygenase in phenylacetate catabolic pathway. This enzyme converts phenylacetyl-CoA to a 2-(1, 2-epoxy-1, 2-dihydrophenyl) acetyl-CoA. Archaea harboring key genes of this pathway are some members of the Halobacteria, which may have acquired a multitude of bacterial genes (Kennedy et al., 2001; Notomista et al., 2003). As shown by our analysis, MRU can degrade pyrogallol and phenylacetate produced by gut microbial in ruminants (Martinez-Fernandez et al., 2017).

Detoxification Systems

MRU operome plays a key role in formaldehyde, inorganic arsenate, and copper detoxification process. It contains 6-phosphogluconate dehydrogenase (*gntZ*) gene as homologous to methanotrophic bacteria such as *Methylophilus methylotrophus* and *Methylobacillus flagellates* (Chistoserdova et al., 2000). The presence of arsenate reductase (*arsC*) and Cu⁺-exporting ATPase (*copA*) provides a defense system to its cells against inorganic arsenate and copper toxicities (Liu et al., 2007).

Macromolecule Modification Systems

MRU operome contains α -2, 3-sialyltransferase gene coding protein having a Rossmann fold with the architecture of the α - β complex. This enzyme catalyzes the transfer of sialic acid from CMP-N-acetyl- β -neuraminic acid to membrane proteins and lipids of the cell wall of MRU (Koga et al., 1993). Dolichyl-phosphate-mannose-protein mannosyltransferase is identified as carbohydrate carriers to transfer mannosyl residues to

the hydroxy group of serine or threonine residues during the post-translational protein modification process of MRU (Podar et al., 2013).

Membrane Transport Systems

We observed 16 HPs contributing to the transport systems of this organism (**Supplementary Table S2**). MRU operome encompasses genes coding for transporter proteins responsible for maintenance of metal homeostasis in particular to magnesium and manganese ions and uptake/export of vitamin, sulfite, and tricarboxylate (Winnen et al., 2003; Weinitschke et al., 2007; Hattori et al., 2007, 2009; Rodionov et al., 2009; Rosch et al., 2009; Mayer et al., 2012; Karpowich et al., 2015). The presence of PurR-regulated permease regulon and Na⁺/H⁺ antiporter protein carries out the exchange Na⁺ for H⁺ across the cytoplasmic membrane of archaea (Rimon et al., 2012). Cell-cell communication and intra-species electron transfer can be mediated by preprotein translocase predicted from its operome, as described for hydrogenotrophic methanogens and *E. coli* (Cooper et al., 2017). Translocation sheath protein has an N-terminal domain that mediates the translocation of SPI-2 TTSS effector proteins in MRU (Nikolaus et al., 2001).

D-Gluconate Catabolic System

As shown by our analysis, we proposed a putative D-gluconate catabolic pathway exclusively present in MRU for the biosynthesis of archaeal membrane phospholipids (**Figure 4**). The presence of six HPs with predicted functions evidences the existence of this pathway in this organism. Klemm et al. (1996), identified a *gntP* gene to be involved in gluconate uptake by *E. coli*. *Haloferax volcanii* contains a DeoR/GlpR-type transcription factor, which has shown its potential role as a global regulator of sugar metabolism and to cotranscribe with the downstream phosphofructokinase (*pfkB*) gene (Rawls et al., 2010). As similar to *Pseudomonas aeruginosa*, MRU operome has D-gluconate kinase gene despite a membrane-bound D-gluconate dehydrogenase gene to synthesize phospholipids (Matsushita et al., 1979; Schlichtman et al., 1995; Kulakova et al., 2001). As similar to archaea, the utilization of gluconate in MRU leads to a branch point for two central metabolic pathways: the Entner-Doudoroff pathway and phospholipids biosynthesis (Bräsen et al., 2014).

DISCUSSION

The function of operome is obscure and quite unsettling in prokaryotic genomes. Understanding important knowledge gaps in the unknown function of operome can unravel their cellular and molecular mechanisms. The functionality of proteins with unknown function have been identified, characterized, and validated with a broad spectrum of genetic and biochemical experiments (Mills et al., 2015). Several computational methods have been used to describe the physiological states of methanogens from the predicted functions of operome (Chellapandi and Prisilla, 2018;

TABLE 3 | Functional annotation of operome involved in different metabolic subsystems of MRU.

Locus tag	Assigned function	EC	Gene
Biosynthesis			
<i>Amino acids biosynthesis</i>			
1696	Carbamoyl-phosphate synthase (glutamine-hydrolyzing)	6.3.5.5	<i>carB</i>
1737	Saccharopine dehydrogenase (NAD/P, L-lysine-forming)	1.5.1.7 1.5.1.8	<i>lys1</i>
Aminoacyl-tRNA charging metabolic clusters			
0488 0490 0499 0764 0780 0790 0801 1131 1147 1150 1364 1590 1796	Methionine—tRNA ligase	6.1.1.10	<i>metG</i>
1493	Tryptophanyl-tRNA synthetase (Membrane bound)	6.1.1.2	<i>trpS</i>
Carbohydrates and Cell structures biosynthesis			
1418	dTDP-4-amino-4,6-dideoxygalactose transaminase	2.6.1.59	<i>rffA</i>
1886	Glycogen Phosphorylase	2.4.1.1	<i>glgP</i>
1469	UDP-glucose 4-epimerase	5.1.3.2	<i>galE</i>
1462	Pantothenate synthase	6.3.2.1	<i>panC</i>
0480	Pyruvate kinase	2.7.1.40	<i>pykA</i>
Cell structures biosynthesis			
1065	CDP-glycerol glycerophosphotransferase	2.7.8.12	<i>tagF</i>
1589 1957	Thiamine monophosphate synthase	2.7.7.39	<i>tagD</i>
Cofactors, Prosthetic groups, Electron carriers biosynthesis			
2219	Cobalamin biosynthesis protein CbiB	6.3.1.10	<i>cbiB</i>
0947	Coenzyme F420-0:L-glutamate ligase	6.3.2.31	<i>cofE</i>
1116	CTP: Molybdenum cofactor cytidyltransferase	2.7.7.76	–
1450	Energy-converting hydrogenase B subunit O	1.6.5.3	<i>ehbO</i>
0776 0785	Gamma-glutamyl cyclotransferase	2.3.2.4	<i>ykqA</i>
1937	Glutathione peroxidase	1.11.1.9	<i>gpxA</i>
0035	NUDIX hydrolase	3.6.1.22	<i>nadM</i>
0596	4-Hydroxybenzoate octaprenyltransferase	2.5.1.39	<i>ubiA</i>
1550	5-Formyltetrahydrofolate cyclo-ligase activity	6.3.3.2	<i>mthfs</i>
1831	Dihydroneopterin aldolase	4.1.2.25	<i>folB</i>
1209	Nicotinate-nucleotide pyrophosphorylase [carboxylating]	2.4.2.19	<i>nadC</i>
0277	NUDIX hydrolase	3.6.1.22	<i>nudC</i>
2172	Riboflavin kinase	2.7.1.161	<i>ribK</i>
0432	Tocopherol cyclase	5.5.1.24	<i>vte1</i>
1728	Phosphomevalonate decarboxylase	4.1.1.99	<i>pmd</i>
Fatty acid and lipid biosynthesis			
0460	Dolichol kinase	2.7.1.108	<i>dolk</i>
1693	Integral Membrane bound Phosphatidate cytidyltransferase	2.7.7.41	<i>cdsA</i>
Metabolic regulators biosynthesis			
0939	6-Phosphofructo-2-kinase Fructose-2,6-bisphosphate 2-phosphatase	2.7.1.105 3.1.3.46	<i>pfkfb3</i>
0393 (K18532 adenylate kinase [EC:2.7.4.3])	Adenylate kinase	2.7.4.3	<i>adk</i>
Nucleosides and nucleotides biosynthesis			
0425	L-Threonylcarbamoyladenylate synthase	2.7.7.87	<i>yrdC/sua5/ywIC</i>
1890	Phosphoribosylaminoimidazole carboxylase	4.1.1.21	<i>purE</i>
0720	Uridylate kinase (DNA binding protein)	2.7.4.22	<i>pyrH</i>
Catabolism			
<i>Alcohols degradation</i>			
0528	Coenzyme B12-dependent diol dehydrase	4.2.1.28	<i>pduC</i>
Amino acids degradation			
2016 0381	Succinylglutamate desuccinylase	3.5.1.96	<i>astE</i>
Aromatic compounds degradation			
1622	4-Carboxymuconolactone decarboxylase	4.1.1.44	<i>pcaC</i>
0476	Phenylacetate-CoA oxygenase	1.14.13.149	<i>paaJ</i>
0313	Pyrogallol hydroxytransferase	1.97.1.2	<i>athL</i>

(Continued)

TABLE 3 | Continued

Locus tag	Assigned function	EC	Gene
C₁ Compounds utilization and assimilation			
2132	Bifunctional formaldehyde-activating enzyme	4.2.1.147/4.1.2.43	<i>fae-hps</i>
1013	Phosphogluconate dehydrogenase (NAD ⁺ -dependent, decarboxylating)	1.1.1.343	<i>gntZ</i>
Inorganic nutrients metabolism			
1280 1936	NADPH-dependent FMN reductase	1.5.1.38	<i>ssuE</i>
0224 0376	Phosphonoacetate hydrolase (membrane bound)	3.11.1.2	<i>phnA</i>
Secondary metabolites degradation			
1330	Carbohydrate kinase (Integral membrane-bound)	2.7.1.4	<i>pfkB</i>
2120	Quercetin dioxygenase	1.13.11.24	<i>qodI</i>
Macromolecule modification			
0421	Alpha-2,3-sialyltransferase	2.4.99.4	<i>siat4a</i>
Small molecule reactions			
1938	Arsenate Reductase (Thioredoxin)	1.20.4.1	<i>arsC</i>
0134	Type I restriction-modification system M subunit HsdM	2.1.1.72	<i>hsdM</i>
0674 1683 1749	Succinate dehydrogenase (quinone)	1.3.5.1	<i>sdh</i>
1013	Phosphogluconate dehydrogenase (NAD ⁺ -dependent, decarboxylating)	1.1.1.343	<i>gntZ</i>
2194	2-Enoyl-CoA Hydratase	3.4.21.92	<i>clpP</i>
0747	2-Polyprenylphenol 6- hydroxylase	1.14.13.-	<i>ubiB2</i>
0202	Aconitate hydratase	4.2.1.3	<i>acnA</i>
2180 2184 2185	Acytransferase	2.3.1.13	<i>glyat</i>
0496	ATP pyrophosphatase	3.6.1.8	<i>thiI</i>
0062 0063 1113 1172	ATP-dependent DNA helicase	3.6.4.12	<i>ashA</i>
2196	Choloylglycine hydrolase	3.5.1.24	–
0156 0041	DNA binding E3 SUMO-protein ligase	6.3.2.-	<i>piaS4</i>
0174 (K09723 DNA replication factor GINS)	DNA primase small subunit	2.7.7.-	<i>priA</i>
2069	DNA-3-methyladenine glycosylase	3.2.2.20	<i>tag</i>
1108 2173	DNA-directed DNA polymerase	2.7.7.7	<i>polB</i>
1660 1699 1734	Flavin reductase	1.5.1.36	<i>hpaC</i>
1442	Geranylgeranyl reductase	1.3.1.83	<i>chlP</i>
1290 1291	Lincosamide nucleotidyltransferase	2.7.7.-	<i>inuA</i>
0930	Manganese-dependent inorganic pyrophosphatase	3.6.1.1	<i>ppaC</i>
0223	Membrane-bound O-acyltransferase	2.3.1.-	<i>rimL</i>
1242	Nucleoside Triphosphate Pyrophosphohydrolase	3.6.1.8	<i>mazG</i>
1605 0049	Nucleotide diphosphatase	3.6.1.9	<i>ENPP</i>
2146	Oligosaccharyl transferase	2.4.99.18	<i>STT3</i>
1588	Succinylglutamate desuccinylase /aspartoacylase	3.5.1.15	<i>aspA</i>
0100	Peptidoglycan-associated polymer biosynthesis	2.-.-.-	<i>csaB</i>
1555	Pseudouridine-5'-monophosphatase	3.1.3.-	<i>HDHD1</i>
1964	Sterol 3-beta-glucosyltransferase (Phosphorylating)	2.4.1.173	–
1631	UDP-N-acetylglucosamine 2-epimerase (non-hydrolyzing)	5.1.3.14	<i>wecB</i>
0835	von Willebrand/Integrin A Domains	3.6.4.-	<i>hepA</i>
Protein-modification reactions			
1344	Lysine carboxypeptidase	3.4.17.3	<i>CPN1</i>
1375	Membrane-bound dolichyl-phosphate-mannose-protein mannosyltransferase	2.4.1.109	<i>pomT</i>
0791	Methylated-DNA—[protein]-cysteine S-methyltransferase	2.1.1.63	–
1884	Nucleotide-activated 6-deoxyhexose biosynthesis	2.4.1.109	<i>pomT</i>
2158	Putative pyruvate formate-lyase	1.97.1.4	<i>pfIX</i>
1801 1867	Ribosomal-protein-alanine N-acetyltransferase	2.3.1.128	<i>rimI</i>
1389 1514	S-Adenosyl-L-methionine-dependent methyltransferase	1.16.1.8	<i>mtrR</i>
1096	Serine/threonine protein kinase with TPR repeats	2.7.11.1	<i>bub1</i>
1563	Proteasome endopeptidase complex	3.4.25.1	<i>psmA</i>
1311 0426	tRNA-splicing ligase	6.5.1.3	<i>rtcB</i>

(Continued)

TABLE 3 | Continued

Locus tag	Assigned function	EC	Gene
Energy metabolism			
Generation of precursor metabolites and energy			
2214	Fuculose 1-phosphate aldolase	4.1.2.17	<i>fucA</i>
1894	Fumarate hydratase	4.2.1.2	<i>fumA</i>

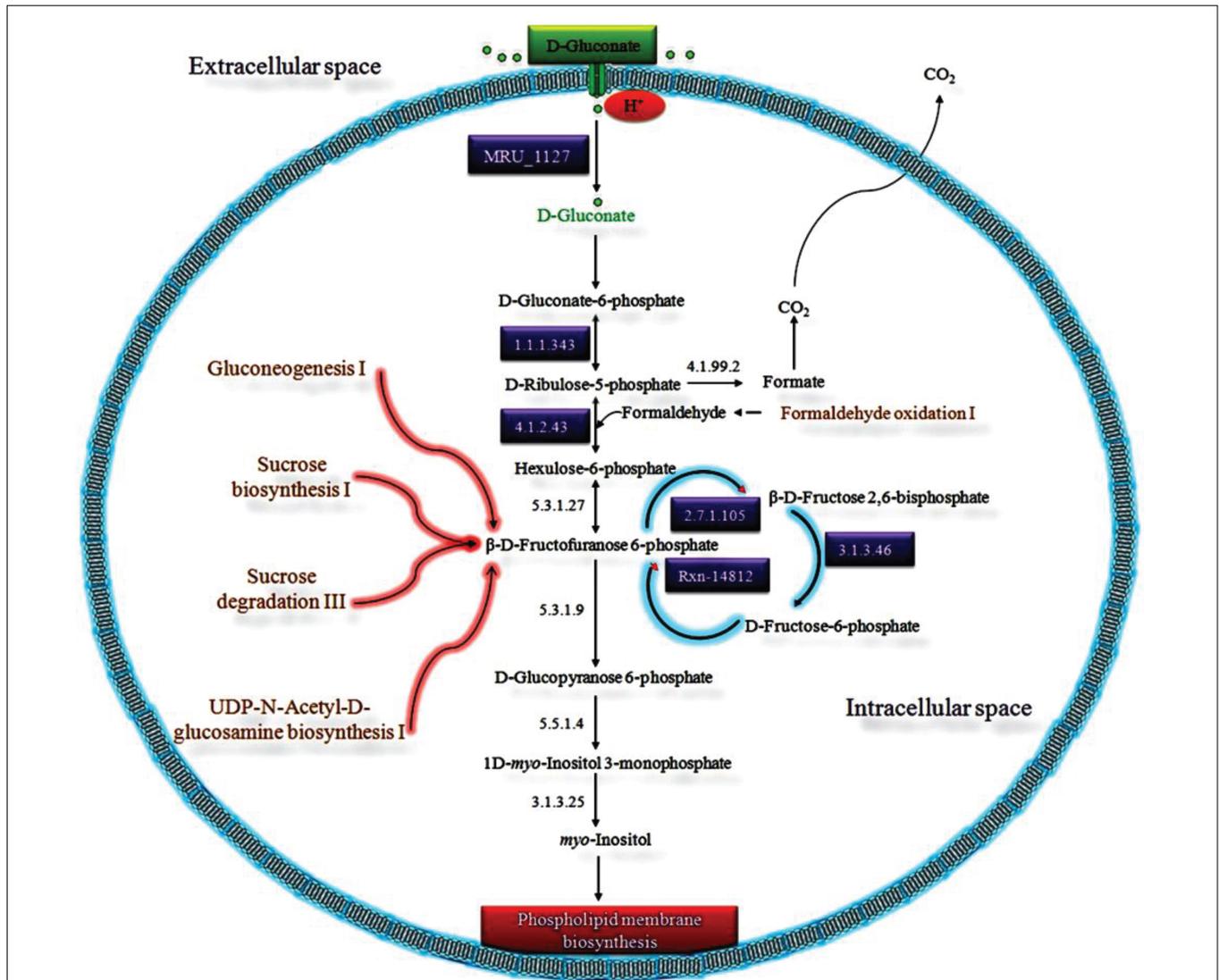


FIGURE 4 | The proposed D-gluconate catabolic pathway in MRU was discovered from the functional annotation of its operome. D-Gluconate is imported into the cytoplasm by the predicted gluconate transporter (*gntP*) gene. It can be phosphorylated to D-gluconate-6-phosphate by D-gluconate kinase (*gntK*), which is then converted to D-ribulose-5-phosphate by the catalytic action of NAD⁺-dependent phosphogluconate dehydrogenase (*gntZ*). D-Ribulose-5-phosphate is next oxidized to hexulose-6-phosphate by 3-hexulose phosphate synthase (*hxlA*) and converted into β-D-fructofuranose 6-phosphate with phospho-3-hexuloisomerase (*phi1*). The 6-phosphofructose 2-kinase phosphorylates β-D-fructofuranose 6-phosphate into β-D-fructose 2,6-bisphosphate, which then interconverts from D-fructose-6-phosphate to β-D-fructofuranose 6-phosphate by fructose-2,6-bisphosphatase. In an alternative way, β-D-fructofuranose 6-phosphate is phosphorylated to D-glucopyranose 6-phosphate by 6-phosphofructo-2-kinase. Glucopyranose 6-phosphate is converted to 1D-myo-inositol 3-monophosphate by D-glucose 6-phosphate cycloaldolase (*ino1*) and reduced to myo-inositol by inositol-phosphate phosphatase (*suhB*).

Prathiviraj and Chellapandi, 2019). There are several functional measures (structural and functional motifs) to be considered for computational predictions of operome from available

microbial genomes. The present study employed to collect comprehensive information derived from sequence similarity, conserved domain, motif, structure, fold, protein-protein

interaction, subcellular localization, phylogenetic inference, and gene expression profile as the predictive measures to assign a precise molecular function to MRU operome. Collective information of them provides a hint to predict some distinct motifs and annotate the function of each protein accurately for studying growth physiology in the rumen ecosystem.

Generally, the protein sequence is less conserved than the tertiary structure of a protein (Illergård et al., 2009). In this study, experimentally solved structures and accurate protein folding offered the major importance to deduce some level of a functional description of a protein, as described by Nealon et al. (2017). Characterization of binding motifs and catalytic cores present in the proteins and functional categorization in the cell has been achieved by using the predictive measures derived from overall proteome information (Shapiro and Harris, 2000). Many protein domains have unknown functions, but they may contribute to the metabolic regulation of organisms (Kotze et al., 2013). It implied the possibility of finding a new domain and motif as well as discovers additional protein pathways and cascades from functionally annotated operome (Ijaq et al., 2015). Functional prediction and assignment of prokaryotic operome have been either only sequence-based or structure-based strategies. In our study, a combination of bioinformatics tools with 6 different prediction schemas and additional literature evidence with a 6-level confidence score was applied to improve the prediction accuracy of our functional assignment (Figure 1). Compared to earlier functional prediction approaches, our approach provides a strong emphasis to reveal its metabolic subsystems and cellular mechanisms from the assigned function of operome.

The mechanisms of molecular pathogenesis and virulence of many pathogenic organisms and drug targets discovery are being considered an accurate prediction of operome function as an important biological knowledgebase (Amavisit et al., 2003; Lamarche et al., 2008; Kumar et al., 2014). Several bioinformatics tools have been utilized for functional prediction of operome from different pathogenic organisms (Kumar et al., 2014, 2015; Singh et al., 2017; Shrivastava et al., 2017). It clearly described that all of them are pathogenic organisms but no reports on rumen methanogens yet. It was the first computational study to characterize the function of MRU operome, a potential methanogen for enteric methane emission in the ruminants via enteric fermentation.

The Rossmann was a novel and ancient fold found in 5, 10-methenyltetrahydromethanopterin hydrogenase, a key enzyme of hydrogenotrophic methanogenesis. It explains the possibility of hydrogenotrophic lifestyle in MRU, as described by Leahy et al. (2010). The reduction potentials of rubredoxin fold-containing proteins are known to be involved in biochemical processes including carbon fixation, detoxification, and fatty acid metabolism (Prakash et al., 2018; Prathviraj and Chellapandi, 2020). Cofactors or other prosthetic groups are more attractive to stimulate enzyme activity in hydrolytic reactions of archaea. Transmembrane helices are generally independently stable in a membrane or membrane-like environment, which are

important for signal recognition, transport phenomena, energy translocation, and conservation in the living cell (von Heijne, 1988; Jennings, 1989). Concerning the functional importance, we classified and categorized the function of MRU operome in this study.

In this study, six operon-like clusters were identified from MRU operome. The functions of predicted gene clusters were contributed in chaperone-assisted protein folding, host defense response, and toxicity neutralization of MRU. Some transcriptional regulatory systems predicted from its operome have shown to protect cellular proteins against sudden heat-shock stress, nitrogen limitation, and heavy metal homeostasis. MRU genome contains many pathway holes, which hinder its accurate metabolic reconstruction at the genome-scale. In our study, we detected some key genes missing in the metabolic network of this organism. Consequently, complete metabolic subsystems were annotated for the biosynthesis of L-lysine, L-glutamate, enterobacterial-like common antigen, archaeal-specific phospholipids, and isopentenyl diphosphate. MRU operome can produce coenzyme F₄₂₀ and flavin and electron carriers. Cell wall lipids and membrane proteins have been synthesized from the function of some HPs through macromolecule modification reactions. This organism has well-established transporter systems to maintain metal homeostasis and uptake/export of vitamin, gluconate, sulfite, and tricarboxylate. D-Gluconate catabolic pathway was uniquely discovered from MRU operome for the biosynthesis of archaeal membrane phospholipids.

CONCLUSION

The functional assignment of operome is a mandatory process for a better understanding of the metabolic and molecular processes of this organism. The predicted functional properties of its operome afford us not only for new structural information but also for new molecular functions essential for the lifestyle in the rumen ecosystem. A major operome covers all functional counterparts needed to perform diverse metabolic pathways and regulatory processes. Some imperative physiological functions (oxidative stress, archaeal-specific membrane phospholipids, etc.) of this organism are revealed from this study. The genome-neighborhood analysis found six main gene clusters (hsp, tram, dsr, cbs, anti-toxin, and gas), which are contributed to the energetic metabolism and defense systems. MRU operome contains 119 metabolic enzymes with 18 sub-pathways and 25 binding proteins that recognize the DNA, RNA, metal, and membrane for cellular function. Interestingly, we discovered a putative D-gluconate catabolic pathway for the biosynthesis of archaeal-specific membrane phospholipids. Several virulence-associated and vaccine targeted proteins have been identified from MRU operome. It suggests the development of new methane mitigation interventions that target the key metabolic proteins to reduce methane emissions in ruminants. Functional prediction and assignment of its operome are thus very important to comprehend the cellular machinery at the systems-level for

anti-methanogenic compounds discovery. Nevertheless, all of our predicted functions of its operome should be evaluated and validated experimentally with protein expression and purification, crystallization, and structure determination studies.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding author/s.

AUTHOR CONTRIBUTIONS

PC: research design, concept, and manuscript writing. MB: data preparation and analysis. NS: data analysis and manuscript revision. All authors contributed to the article and approved the submitted version.

REFERENCES

- Agnihotri, G., and Liu, H. W. (2003). Enoyl-CoA hydratase. reaction, mechanism, and inhibition. *Bioorg. Med. Chem.* 11, 9–20.
- Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W., et al. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 25, 3389–3402. doi: 10.1093/nar/25.17.3389
- Amavisit, P., Lightfoot, D., Browning, G. F., and Markham, P. F. (2003). Variation between pathogenic serovars within *salmonella* pathogenicity islands. *J. Bacteriol.* 185, 3624–3635. doi: 10.1128/jb.185.12.3624-3635.2003
- Barbas, A., Popescu, A., Frazão, C., Arraiano, C. M., and Fialho, A. M. (2013). Rossmann-fold motifs can confer multiple functions to metabolic enzymes: RNA binding and ribonuclease activity of a UDP-glucose dehydrogenase. *Biochem. Biophys. Res. Commun.* 430, 218–224. doi: 10.1016/j.bbrc.2012.10.091
- Benaroudj, N., and Goldberg, A. L. (2000). PAN, the proteasome-activating nucleotidase from archaeobacteria, is a protein-unfolding molecular chaperone. *Nat. Cell Biol.* 2, 833–839. doi: 10.1038/35041081
- Benkert, P., Tosatto, S. C., and Schomburg, D. (2008). QMEAN: a comprehensive scoring function for model quality assessment. *Proteins* 71, 261–277. doi: 10.1002/prot.21715
- Bharat Siva Varma, P., Adimulam, Y. B., and Kodukula, S. (2015). Insilico functional annotation of a hypothetical protein from *Staphylococcus aureus*. *J. Infect. Publ. Health* 8, 526–532. doi: 10.1016/j.jiph.2015.03.007
- Biasini, M., Bienert, S., Waterhouse, A., Arnold, K., Studer, G., Schmidt, T., et al. (2014). SWISS-MODEL: modeling protein tertiary and quaternary structure using evolutionary information. *Nucleic Acids Res.* 42, W252–W258.
- Bräsen, C., Esser, D., Rauch, B., and Siebers, B. (2014). Carbohydrate metabolism in Archaea: current insights into unusual enzymes and pathways and their regulation. *Microbiol. Mol. Biol. Rev.* 78, 89–175. doi: 10.1128/mmr.00041-13
- Carter, J. R., Fox, C. F., and Kennedy, E. P. (1968). Interaction of sugars with the membrane protein component of the lactose transport system of *Escherichia coli*. *Proc. Natl. Acad. Sci. U.S.A.* 60, 725–732. doi: 10.1073/pnas.60.2.725
- Caspi, R., Altman, T., Billington, R., et al. (2014). The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of Pathway/Genome Databases. *Nucleic Acids Res.* 42, D459–D471.
- Chang, F. M., Coyne, H. J., Cubillas, C., Vinuesa, P., Fang, X., Ma, Z., et al. (2014). Cu(I)-mediated allosteric switching in a copper-sensing operon repressor (*CsoR*). *J. Biol. Chem.* 289, 19204–19217. doi: 10.1074/jbc.m114.556704
- Chellapandi, P., Bharathi, M., Sangavai, C., and Prathiviraj, R. (2018). *Methanobacterium formicicum* as a target rumen methanogen for the development of new methane mitigation interventions-A review. *Veter. Anim. Sci.* 6, 86–94. doi: 10.1016/j.vas.2018.09.001

ACKNOWLEDGMENTS

We would like to thank the University Grants Commission (RA-2012-14-SC-TAM-1768) and Department of Biotechnology (BT/49/NE/2014), New Delhi, India for financial assistance.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2020.593990/full#supplementary-material>

Supplementary Table 1 | Functional annotation of operome involved in diverse metabolic systems of MRU.

Supplementary Table 2 | Functional annotation of operome involved in transporter mechanisms of MRU.

Supplementary Data | All predicted information for functional assignment of MRU operome.

- Chellapandi, P., Bharathi, M., Prathiviraj, R., Sasikala, R., and Vikraman, M. (2017a). Genome-scale metabolic model as a virtual platform to reveal the ecological importance of methanogenic archaea. *Curr. Biotechnol.* 6, 149–160. doi: 10.2174/2211550105666160901125353
- Chellapandi, P., Mohamed Khaja, Hussain, M., and Prathiviraj, R. (2017b). CPSIR-CM: a database for structural properties of proteins identified in cyanobacterial C1 metabolism. *Algal. Res.* 22, 135–139. doi: 10.1016/j.algal.2016.12.005
- Chellapandi, P., and Prathiviraj, R. (2020). A systems biology perspective of *Methanothermobacter thermoautotrophicus* strain ΔH for bioconversion of CO₂ to methane. *J. CO₂ Utiliz.* 40:101210. doi: 10.1016/j.jcou.2020.101210
- Chellapandi, P., and Prisilla, A. (2018). *Clostridium botulinum* type A-virulome-gut interactions: a systems biology insight. *Hum. Microb. J.* 8, 15–22. doi: 10.1016/j.humic.2018.01.003
- Chellapandi, P., and Ranjani, J. (2015). Knowledge-based discovery for designing CRISPR-CAS systems against invading mobilomes in thermophiles. *Syst. Synth. Biol.* 9, 97–106. doi: 10.1007/s11693-015-9176-8
- Cheng, A. G., McAadow, M., Kim, H. K., Bae, T., Missiakas, D. M., and Schneewind, O. (2010). Contribution of coagulases towards *Staphylococcus aureus* disease and protective immunity. *PLoS Pathog.* 6:e1001036. doi: 10.1371/journal.ppat.1001036
- Chistoserdova, L., Gomelsky, L., Vorholt, J. A., Gomelsky, M., Tsygankov, Y. D., and Lidstrom, M. E. (2000). Analysis of two formaldehyde oxidation pathways in *Methylobacillus flagellatus* KT, a ribulose monophosphate cycle methylotroph. *Microbiology* 146, 233–238. doi: 10.1099/00221287-146-1-233
- Cooper, R. M., Tsimring, L., and Hasty, J. (2017). Inter-species population dynamics enhance microbial horizontal gene transfer and spread of antibiotic resistance. *eLife* 6:e25950.
- da Fonseca, M. M., Zaha, A., Caffarena, E. R., and Vasconcelos, A. T. (2012). Structure-based functional inference of hypothetical proteins from *Mycoplasma hyopneumoniae*. *J. Mol. Model* 18, 1917–1925. doi: 10.1007/s00894-011-1212-3
- de Castro, E., Sigrist, C. J., Gattiker, A., Bulliard, V., Langendijk-Genevaux, P. S., Gasteiger, E., et al. (2006). ScanProsite: detection of PROSITE signature matches and ProRule-associated functional and structural residues in proteins. *Nucleic Acids Res.* 34, W362–W365.
- Doerks, T., van Noort, V., Minguez, P., and Bork, P. (2012). Annotation of the *M. tuberculosis* hypothetical orfeome: adding functional information to more than half of the uncharacterized proteins. *PLoS One* 7:e34302. doi: 10.1371/journal.pone.0034302
- Dokland, T. (1999). Scaffolding proteins and their role in viral assembly. *Cell Mol. Life Sci.* 56, 580–603. doi: 10.1007/s000180050455
- Enzmann, F., Mayer, F., Rother, M., and Holtmann, D. (2018). Methanogens: biochemical background and biotechnological applications. *AMB Exp.* 8:1.

- Fay, A. W., Wiig, J. A., Lee, C. C., and Hu, Y. (2015). Identification and characterization of functional homologs of nitrogenase cofactor biosynthesis protein *NifB* from methanogens. *Proc. Natl. Acad. Sci. U.S.A.* 112, 14829–14833. doi: 10.1073/pnas.1510409112
- Finn, R. D., Attwood, T. K., Babbitt, P. C., et al. (2016). InterPro in 2017—beyond protein family and domain annotations. *Nucleic Acids Res.* 45, D190–D199.
- Forouhar, F., Saadat, N., Hussain, M., Seetharaman, J., Lee, I., Janjua, H., et al. (2011). A large conformational change in the putative ATP pyrophosphatase PF0828 induced by ATP binding. *Acta Crystallogr. Sect. F. Struct. Biol. Cryst. Commun.* 67, 1323–1327. doi: 10.1107/s1744309111031447
- Garfoot, A. L., Zemska, O., and Rappleye, C. A. (2014). *Histoplasma capsulatum* depends on de novo vitamin biosynthesis for intraphagosomal proliferation. *Infect. Immun.* 82, 393–404. doi: 10.1128/iai.00824-13
- Gaywet, J., Xu, W., Radulovic, S., Bessman, M. J., and Azad, A. F. (2002). The Rickettsia prowazekii invasion gene homolog (*invA*) encodes a Nudix hydrolase active on adenosine (5′)-pentaphospho-(5′)-adenosine. *Mol. Cell Proteom.* 1, 179–185. doi: 10.1074/mcp.m100030-mcp200
- Geourjon, C., and Deléage, G. (1995). SOPMA: significant improvements in protein secondary structure prediction by consensus prediction from multiple alignments. *Comput. Appl. Biosci.* 11, 681–684. doi: 10.1093/bioinformatics/11.6.681
- Giaquinto, L., Curmi, P. M., Siddiqui, K. S., Poljak, A., DeLong, E., DasSarma, S., et al. (2007). Structure and function of cold shock proteins in archaea. *J. Bacteriol.* 189, 5738–5748. doi: 10.1128/jb.00395-07
- Greenbaum, D., Luscombe, N. M., Jansen, R., Qian, J., and Gerstein, M. (2001). Interrelating different types of genomic data, from proteome to secretome: 'oming in on function. *Genome Res.* 11, 1463–1468. doi: 10.1101/gr.207401
- Grochowski, L. L., Xu, H., and White, R. H. (2006). *Methanocaldococcus jannaschii* uses a modified mevalonate pathway for biosynthesis of isopentenyl diphosphate. *J. Bacteriol.* 188, 3192–3198. doi: 10.1128/jb.188.9.3192-3198.2006
- Grogan, D. W. (2015). Understanding DNA repair in hyperthermophilic archaea: persistent gaps and other reasons to focus on the fork. *Archaea* 2015;942605.
- Guruprasad, K., Reddy, B. V., and Pandit, M. W. (1990). Correlation between stability of a protein and its dipeptide composition: a novel approach for predicting *in vivo* stability of a protein from its primary sequence. *Prot. Eng.* 4, 155–161. doi: 10.1093/protein/4.2.155
- Hattori, M., Iwase, N., Furuya, N., Tanaka, Y., Tsukazaki, T., Ishitani, R., et al. (2009). Mg²⁺-dependent gating of bacterial *MgtE* channel underlies Mg²⁺ homeostasis. *EMBO J.* 28, 3602–3612. doi: 10.1038/emboj.2009.288
- Hattori, M., Tanaka, Y., Fukai, S., Ishitani, R., and Nureki, O. (2007). Crystal structure of the *MgtE* Mg²⁺ transporter. *Nature* 448, 1072–1075. doi: 10.1038/nature06093
- He, Z. G., Feng, Y., Wang, J., and Jiang, P. X. (2008). The regulatory function of N-terminal AAA+ ATPase domain of eukaryote-like archaeal *Orc1/Cdc6* protein during DNA replication initiation. *Arch. Biochem. Biophys.* 471, 176–183. doi: 10.1016/j.abb.2008.01.007
- Homa, F. L., and Brown, J. C. (1997). Capsid assembly and DNA packaging in herpes simplex virus. *Rev. Med. Virol.* 7, 107–122. doi: 10.1002/(sici)1099-1654(199707)7:2<107::aid-rmv191>3.0.co;2-m
- Hristov, A. N., Oh, J., Firkins, J. L., Dijkstra, J., Kebreab, E., Waghorn, G., et al. (2013). Special topics—Mitigation of methane and nitrous oxide emissions from animal operations: I. A review of enteric methane mitigation options. *J. Anim. Sci.* 91, 5045–5069. doi: 10.2527/jas.2013-6583
- Hubbard, K. E., and Dunbar, S. D. (2017). Perceptions of scientific research literature and strategies for reading papers depend on academic career stage. *PLoS One* 12:e0189753. doi: 10.1371/journal.pone.0189753
- Hwang, B. Y., Lee, H. J., Yang, Y. H., Joo, H. S., and Kim, B. G. (2004). Characterization and investigation of substrate specificity of the sugar aminotransferase *WecE* from *E. coli* K12. *Chem. Biol.* 11, 915–925. doi: 10.1016/j.chembiol.2004.04.015
- Ijaq, J., Chandrasekharan, M., Poddar, R., Bethi, N., and Sundararajan, V. S. (2015). Annotation and curation of uncharacterized proteins—challenges. *Front. Genet.* 6:119. doi: 10.3389/fgene.2015.00119
- Ikai, A. (1980). Thermostability and aliphatic index of globular proteins. *J. Biochem.* 88, 1895–1898.
- Illergård, K., Ardell, D. H., and Elofsson, A. (2009). Structure is three to ten times more conserved than sequence—a study of structural response in protein cores. *Proteins* 77, 499–508. doi: 10.1002/prot.22458
- Islam, M. S., Shahik, S. M., Soheli, M., Patwary, N. I., and Hasan, M. A. (2015). *In silico* structural and functional annotation of hypothetical proteins of *Vibrio cholerae* O139. *Genom. Inform.* 13, 53–59.
- Janssen, P. H., and Kirs, M. (2008). Structure of the archaeal community of the rumen. *Appl. Environ. Microbiol.* 74, 3619–3625. doi: 10.1128/aem.02812-07
- Jennings, M. L. (1989). Topography of membrane proteins. *Annu. Rev. Biochem.* 58, 999–1027. doi: 10.1146/annurev.bi.58.070189.005031
- Jitendra, S., Narula, R., Agnihotri, S., and Singh, M. (2011). Annotation of hypothetical proteins orthologous in *Pongo abelii* and *Sus scrofa*. *Bioinformatics* 6, 297–299. doi: 10.6026/97320630006297
- Kallioma-Sanford, A. K., Rodriguez-Castañeda, F. A., McLeod, B. N., Latorre-Roselló, V., Smith, J. H., Reimann, J., et al. (2012). Chromosome segregation in Archaea mediated by a hybrid DNA partition machine. *Proc. Natl. Acad. Sci. U.S.A.* 109, 3754–3759. doi: 10.1073/pnas.1113384109
- Kandiba, L., Guan, Z., and Eichler, J. (2013). Lipid modification gives rise to two distinct *Haloferax volcanii* S-layer glycoprotein populations. *Biochim. Biophys. Acta* 1828, 938–943. doi: 10.1016/j.bbamem.2012.11.023
- Kanehisa, M., Sato, Y., Furumichi, M., Morishima, K., and Tanabe, M. (2018). New approach for understanding genome variations in KEGG. *Nucleic Acids Res.* 47, D590–D595. doi: 10.1093/nar/gky962
- Karpowich, N. K., Song, J. M., Cocco, N., and Wang, D. N. (2015). ATP binding drives substrate capture in an ECF transporter by a release-and-catch mechanism. *Nat. Struct. Mol. Biol.* 22, 565–571. doi: 10.1038/nsmb.3040
- Kaster, A. K., Moll, J., Parey, K., and Thauer, R. K. (2011). Coupling of ferredoxin and heterodisulfide reduction via electron bifurcation in hydrogenotrophic methanogenic archaea. *Proc. Natl. Acad. Sci. USA.* 108, 2981–2986. doi: 10.1073/pnas.1016761108
- Kennedy, S. P., Ng, W. V., Salzberg, S. L., Hood, L., and DasSarma, S. (2001). Understanding the adaptation of Halobacterium species NRC-1 to its extreme environment through computational analysis of its genome sequence. *Genome Res.* 11, 1641–1650. doi: 10.1101/gr.190201
- Klemm, P., Tong, S., Nielsen, H., and Conway, T. (1996). The *thegntP* gene of *Escherichia coli* involved in gluconate uptake. *J. Bacteriol.* 178, 61–67. doi: 10.1128/jb.178.1.61-67.1996
- Knudsen, M., and Wiuf, C. (2010). The CATH database. *Hum. Genom.* 4, 207–212. doi: 10.1186/1479-7364-4-3-207
- Koga, Y., Nishihara, M., Morii, H., and Akagawa-Matsushita, M. (1993). Ether polar lipids of methanogenic bacteria: structures, comparative aspects, and biosyntheses. *Microbiol. Rev.* 57, 164–182. doi: 10.1128/mmr.57.1.164-182.1993
- Kotze, H. L., Armitage, E. G., Sharkey, K. J., Allwood, J. W., Dunn, W. B., Williams, K. J., et al. (2013). A novel untargeted metabolomics correlation-based network analysis incorporating human metabolic reconstructions. *BMC Syst. Biol.* 7:107. doi: 10.1186/1752-0509-7-107
- Krogh, A., Larsson, B., von Heijne, G., and Sonnhammer, E. L. (2001). Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J. Mol. Biol.* 305, 567–580. doi: 10.1006/jmbi.2000.4315
- Kulakova, A. N., Kulakov, L. A., Akulenko, N. V., Ksenzenko, V. N., Hamilton, J. T., and Quinn, J. P. (2001). Structural and functional analysis of the phosphonoacetate hydrolase (*phnA*) gene region in *Pseudomonas fluorescens* 23F. *J. Bacteriol.* 183, 3268–3275. doi: 10.1128/jb.183.11.3268-3275.2001
- Kumar, K., Prakash, A., Anjum, F., Islam, A., Ahmad, F., and Hassan, M. I. (2015). Structure-based functional annotation of hypothetical proteins from *Candida dubliniensis*: a quest for potential drug targets. *3 Biotech.* 5, 561–576. doi: 10.1007/s13205-014-0256-3
- Kumar, K., Prakash, A., Tasleem, M., Islam, A., Ahmad, F., and Hassan, M. I. (2014). Functional annotation of putative hypothetical proteins from *Candida dubliniensis*. *Gene* 543, 93–100. doi: 10.1016/j.gene.2014.03.060
- Kyte, J., and Doolittle, R. F. (1982). A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.* 157, 105–132. doi: 10.1016/0022-2836(82)90515-0
- Lamarche, M. G., Wanner, B. L., Crépin, S., and Harel, J. (2008). The phosphate regulon and bacterial virulence: a regulatory network connecting phosphate homeostasis and pathogenesis. *FEMS Microbiol. Rev.* 32, 461–473. doi: 10.1111/j.1574-6976.2008.00101.x
- Large, A. T., Goldberg, M. D., and Lund, P. A. (2009). Chaperones and protein folding in the archaea. *Biochem. Soc. Trans.* 37, 46–51. doi: 10.1042/bst0370046

- Le, N. Q. K., and Huynh, T. T. (2019). Identifying SNAREs by Incorporating Deep Learning Architecture and Amino Acid Embedding Representation. *Front. Physiol.*10:1501. doi: 10.3389/fphys.2019.01501
- Le, N. Q. K., Huynh, T. T., Yapp, E. K. Y., and Yeh, H. Y. (2019). Identification of clathrin proteins by incorporating hyperparameter optimization in deep learning and PSSM profiles. *Comput. Methods Progr. Biomed.*177, 81–88. doi: 10.1016/j.cmpb.2019.05.016
- Leahy, S. C., Kelly, W. J., Altermann, E., Ronimus, R. S., Yeoman, C. J., Pacheco, D. M., et al. (2010). The genome sequence of the rumen methanogen *Methanobrevibacter ruminantium* reveals new possibilities for controlling ruminant methane emissions. *PLoS One*5:e8926. doi: 10.1371/journal.pone.0008926
- Letunic, I., Doerks, T., and Bork, P. (2012). SMART 7: recent updates to the protein domain annotation resource. *Nucleic Acids Res.*40, D302–D305.
- Li, H., Xu, H., Graham, D. E., and White, R. H. (2003). Glutathione synthetase homologs encode alpha-L-glutamate ligases for methanogenic coenzyme F₄₂₀ and tetrahydroscarinapterin biosyntheses. *Proc. Natl. Acad. Sci. U.S.A.*100, 9785–9790. doi: 10.1073/pnas.1733391100
- Liu, J., Huang, C., Shin, D. H., Yokota, H., Jancaric, J., Kim, J. S., et al. (2005). Crystal structure of a heat-inducible transcriptional repressor *HrcA* from *Thermotoga maritima*: structural insight into DNA binding and dimerization. *J. Mol. Biol.*350, 987–996. doi: 10.1016/j.jmb.2005.04.021
- Liu, T., Ramesh, A., Ma, Z., Ward, S. K., Zhang, L., George, G. N., et al. (2007). *CsoR* is a novel *Mycobacterium tuberculosis* copper-sensing transcriptional regulator. *Nat. Chem. Biol.*3, 60–68. doi: 10.1038/nchembio844
- Loewenstein, Y., Raimondo, D., Redfern, O. C., Watson, J., Frishman, D., Linal, M., et al. (2009). Protein function annotation by homology-based inference. *Genome Biol.*10:207. doi: 10.1186/gb-2009-10-2-207
- Louwen, R., Staals, R. H. J., Endtz, H. P., van Baarlen, P., and van der Oost, J. (2014). The role of CRISPR-cas systems in virulence of pathogenic bacteria. *Microbiol. Mol. Biol.*78, 74–88. doi: 10.1128/mmr.00039-13
- Major, T. A., Liu, Y., and Whitman, W. B. (2010). Characterization of energy-conserving hydrogenase B in *Methanococcus maripaludis*. *J. Bacteriol.*192, 4022–4030. doi: 10.1128/jb.101446-09
- Makarova, K. S., Wolf, Y. I., van der Oost, J., and Koonin, E. V. (2009). Prokaryotic homologs of Argonaute proteins are predicted to function as key components of a novel system of defense against mobile genetic elements. *Biol. Direct.*4:29. doi: 10.1186/1745-6150-4-29
- Mao, C., Shukla, M., Larrouy-Maumus, G., Dix, F. L., Kelley, L. A., Sternberg, M. J., et al. (2013). Functional assignment of *Mycobacterium tuberculosis* proteome revealed by genome-scale fold-recognition. *Tuberculosis*93, 40–46. doi: 10.1016/j.tube.2012.11.008
- Marchler-Bauer, A., Derbyshire, M. K., Gonzales, N. R., Lu, S., Chitsaz, F., Geer, L. Y., et al. (2015). CDD: NCBI's conserved domain database. *Nucleic Acids Res.*43, D222–D226.
- Martinez-Fernandez, G., Denman, S. E., Cheung, J., and McSweeney, C. S. (2017). Phloroglucinol degradation in the rumen promotes the capture of excess hydrogen generated from methanogenesis inhibition. *Front. Microbiol.*8:1871. doi: 10.3389/fmicb.2017.01871
- Matsushita, K., Shinagawa, E., Adachi, O., and Ameyama, M. (1979). Membrane-bound D-gluconate dehydrogenase from *Pseudomonas aeruginosa*. Purification and structure of cytochrome-binding form. *J. Biochem.*85, 1173–1181.
- Maupin-Furlow, J. A. (2013). Ubiquitin-like proteins and their roles in archaea. *Trends Microbiol.*21, 31–38. doi: 10.1016/j.tim.2012.09.006
- Mayer, J., Denger, K., Hollemeyer, K., Schleheck, D., and Cook, A. M. (2012). (R)-Cysteate-nitrogen assimilation by *Cupriavidus necator* H16 with excretion of 3-sulfolactate: a patchwork pathway. *Arch. Microbiol.*194, 949–957. doi: 10.1007/s00203-012-0825-y
- Mazandu, G. K., and Mulder, N. J. (2012). Function prediction and analysis of *Mycobacterium tuberculosis* hypothetical proteins. *Int. J. Mol. Sci.*13, 7283–7302. doi: 10.3390/ijms13067283
- McAdow, M., DeDent, A. C., Emolo, C., Cheng, A. G., Kreiswirth, B. N., Missiakas, D. M., et al. (2012). Coagulases as determinants of protective immune responses against *Staphylococcus aureus*. *Infect. Immun.*80, 3389–3398. doi: 10.1128/iai.00562-12
- McAdow, M., Kim, H. K., Dedent, A. C., Hendrickx, A. P., Schneewind, O., and Missiakas, D. M. (2011). Preventing *Staphylococcus aureus* sepsis through the inhibition of its agglutination in blood. *PLoS Pathog.*7:e1002307. doi: 10.1371/journal.ppat.1002307
- Meier-Dieter, U., Starman, R., Barr, K., Mayer, H., and Rick, P. D. (1990). Biosynthesis of enterobacterial common antigen in *Escherichia coli*. Biochemical characterization of Tn10 insertion mutants defective in enterobacterial common antigen synthesis. *J. Biol. Chem.*265, 13490–13497.
- Mills, C. L., Beuning, P. J., and Ondrechen, M. J. (2015). Biochemical functional predictions for protein structures of unknown or uncertain function. *Comput. Struct. Biotechnol. J.*13, 182–191. doi: 10.1016/j.csbj.2015.02.003
- Mitaku, S., Hirokawa, T., and Tsuji, T. (2002). Amphiphilicity index of polar amino acids as an aid in the characterization of amino acid preference at membrane-water interfaces. *Bioinformatics*18, 608–616. doi: 10.1093/bioinformatics/18.4.608
- Namboori, S., Mhatre, N., Sujatha, S., Srinivasan, N., and Pandit, S. B. (2004). Enhanced functional and structural domain assignments using remote similarity detection procedures for proteins encoded in the genome of *Mycobacterium tuberculosis* H37Rv. *J. Biosci.*29, 245–259. doi: 10.1007/bf02702607
- Nealon, J. O., Philomina, L. S., and McGuffin, L. J. (2017). Predictive and experimental approaches for elucidating protein-protein interactions and quaternary structures. *Int. J. Mol. Sci.*18:E2623.
- Nikolaus, T., Deiwick, J., Rappl, C., Freeman, J. A., Schroder, W., Miller, S. I., et al. (2001). SseBCD proteins are secreted by the type III secretion system of *Salmonella* pathogenicity island 2 and function as a translocon. *J. Bacteriol.*183, 6036–6045. doi: 10.1128/jb.183.20.6036-6045.2001
- Notomista, E., Lahm, A., Di Donato, A., and Tramontano, A. (2003). Evolution of bacterial and archaeal multicopon monooxygenases. *J. Mol. Evol.*56, 435–445. doi: 10.1007/s00239-002-2414-1
- Petersen, T. N., Brunak, S., von Heijne, G., and Nielsen, H. (2011). SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat. Methods*8, 785–786. doi: 10.1038/nmeth.1701
- Podar, M., Makarova, K. S., Graham, D. E., Wolf, Y. I., Koonin, E. V., and Reysenbach, A. L. (2013). Insights into archaeal evolution and symbiosis from the genomes of a nanoarchaeon and its inferred crenarchaeal host from Obsidian Pool, Yellowstone National Park. *Biol. Direct.*8:9.
- Poulsen, C., Akhter, Y., Jeon, A. H., Schmitt-Ulms, G., Meyer, H. E., Stefanski, A., et al. (2010). Proteome-wide identification of mycobacterial pupylation targets. *Mol. Syst. Biol.*6:386. doi: 10.1038/msb.2010.39
- Prakash, D., Walters, K. A., Martinie, R. J., McCarver, A. C., Kumar, A. K., Lessner, D. J., et al. (2018). Toward a mechanistic and physiological understanding of a ferredoxin:disulfide reductase from the domains archaea and bacteria. *J. Biol. Chem.*293, 9198–9209. doi: 10.1074/jbc.ra118.002473
- Prathiviraj, R., and Chellapandi, P. (2019). Functional annotation of operome from *Methanothermobacter thermoautotrophicus* ΔH: An insight to metabolic gap filling. *Int. J. Biol. Macromol.*123, 350–362. doi: 10.1016/j.ijbiomac.2018.11.100
- Prathiviraj, R., and Chellapandi, P. (2020). Comparative genomic analysis reveals starvation survival systems in *Methanothermobacter thermoautotrophicus* ΔH. *Anaerobe*64:102216. doi: 10.1016/j.anaerobe.2020.102216
- Prathiviraj, R., and Chellapandi, P. (2020a). Comparative genomic analysis reveals starvation survival systems in *Methanothermobacter thermoautotrophicus* ΔH. *Anaerobe*64:102216.
- Prathiviraj, R., and Chellapandi, P. (2020b). Modeling a global regulatory network of *Methanothermobacter thermoautotrophicus* strain ΔH. *Netw. Model. Anal. Health Inform. Bioinform.*9:17.
- Pulendran, B., and Ahmed, R. (2006). Translating innate immunity into immunological memory: implications for vaccine development. *Cell*124, 849–863. doi: 10.1016/j.cell.2006.02.019
- Quevillon, E., Silventoinen, V., Pillai, S., Harte, N., Mulder, N., Apweiler, R., et al. (2005). InterProScan: protein domains identifier. *Nucleic Acids Res.*33, W116–W120.
- Radivojac, P., Clark, W. T., Oron, T. R., Schnoes, A. M., Wittkop, T., Sokolov, A., et al. (2013). A large-scale evaluation of computational protein function prediction. *Nat. Methods*10, 221–227.
- Rath, V. L., Ammirati, M., LeMotte, P. K., Fennell, K. F., Mansour, M. N., Danley, D. E., et al. (2000). Activation of human liver glycogen phosphorylase by alteration of the secondary structure and packing of the catalytic core. *Mol. Cell.*6, 139–148. doi: 10.1016/s1097-2765(05)00006-7

- Rawls, K. S., Yacovone, S. K., and Maupin-Furrow, J. A. (2010). GlpR represses fructose and glucose metabolic enzymes at the level of transcription in the haloarchaeon *Haloferax volcanii*. *J. Bacteriol.* 192, 6251–6260. doi: 10.1128/jb.00827-10
- Rimon, A., Kozachkov-Magrisso, L., and Padan, E. (2012). The unwound portion dividing helix IV of NhaA undergoes a conformational change at physiological pH and lines the cation passage. *Biochemistry* 51, 9560–9569. doi: 10.1021/bi301030x
- Rodionov, D. A., Hebbeln, P., Eudes, A., ter Beek, J., Rodionova, I. A., Erkens, G. B., et al. (2009). A novel class of modular transporters for vitamins in prokaryotes. *J. Bacteriol.* 191, 42–51. doi: 10.1128/jb.01208-08
- Rosch, J. W., Gao, G., Ridout, G., Wang, Y. D., and Tuomanen, E. I. (2009). Role of the manganese efflux system *mntE* for signalling and pathogenesis in *Streptococcus pneumoniae*: roles of *algR2* and *algH*. *J. Bacteriol.* 177, 2469–2474. doi: 10.1128/jb.177.9.2469-2474.1995
- Sahraeian, S. M., Luo, K. R., and Brenner, S. E. (2015). SIFTER search: a web server for accurate phylogeny-based protein function prediction. *Nucleic Acids Res.* 43, W141–W147.
- Sangavai, C., Prathiviraj, R., and Chellapandi, P. (2020). Functional prediction, characterization and categorization of operome from *Acetoanaerobium sticklandii* DSM 519. *Anaerobe* 61:102088. doi: 10.1016/j.anaerobe.2019.102088
- Schlichtman, D., Kubo, M., Shankar, S., and Chakrabarty, A. M. (1995). Regulation of nucleoside diphosphate kinase and secreted virulence factors in *Pseudomonas aeruginosa*: roles of *algR2* and *algH*. *J. Bacteriol.* 177, 2469–2474. doi: 10.1128/jb.177.9.2469-2474.1995
- Shahbaaz, M., Hassan, M. I., and Ahmad, F. (2013). Functional annotation of conserved hypothetical proteins from *Haemophilus influenzae* Rd KW20. *PLoS One* 8:e84263. doi: 10.1371/journal.pone.0084263
- Shapiro, L., and Harris, T. (2000). Finding function through structural genomics. *Curr. Opin. Biotechnol.* 11, 31–35. doi: 10.1016/s0958-1669(99)00064-6
- Shrivastava, A. K., Kumar, S., Sahu, P. S., and Mahapatra, R. K. (2017). *In silico* identification and validation of a novel hypothetical protein in *Cryptosporidium hominis* and virtual screening of inhibitors as therapeutics. *Parasitol Res.* 116, 1533–1544. doi: 10.1007/s00436-017-5430-1
- Singh, G., and Singh, V. (2018). Functional elucidation of hypothetical proteins for their indispensable roles toward drug designing targets from *Helicobacter pylori* strain HPAG1. *J. Biomol. Struct. Dyn.* 1:13.
- Singh, S., Singh, S. K., Chowdhury, I., and Singh, R. (2017). Understanding the mechanism of bacterial biofilms resistance to antimicrobial agents. *Open Microbiol. J.* 11, 53–62. doi: 10.2174/1874285801711010053
- Sivashankari, S., and Shanmughavel, P. (2006). Functional annotation of hypothetical proteins - A review. *Bioinformation* 1, 335–338. doi: 10.6026/97320630001335
- Terstappen, G. C., and Reggiani, A. (2001). *In silico* research in drug discovery. *Trends Pharmacol. Sci.* 22, 23–26.
- Thieringer, H. A., Jones, P. G., and Inouye, M. (1998). Cold shock and adaptation. *Bioessays* 20, 49–57. doi: 10.1002/(sici)1521-1878(199801)20:1<49::aid-bies8>3.0.co;2-n
- Thompson, J. D., Gibson, T. J., and Higgins, D. G. (2002). Multiple sequence alignment using ClustalW and ClustalX. *Curr. Protoc. Bioinform.* 2, 2.3.
- Tusnády, G. E., and Simon, I. (2001). The HMMTOP transmembrane topology prediction server. *Bioinformatics* 17, 849–850. doi: 10.1093/bioinformatics/17.9.849
- Unterholzner, S. J., Poppenberger, B., and Rozhon, W. (2013). Toxin-antitoxin systems: Biology, identification, and application. *Mob. Genet. Elements* 3:e26219. doi: 10.4161/mge.26219
- Vershon, A. K., Bowie, J. U., Karplus, T. M., and Sauer, R. T. (1986). Isolation and analysis of arc repressor mutants: evidence for an unusual mechanism of DNA binding. *Proteins* 1, 302–311. doi: 10.1002/prot.340010404
- von Heijne, G. (1988). Transcending the impenetrable: how proteins come to terms with membranes. *Biochim. Biophys. Acta* 947, 307–333. doi: 10.1016/0304-4157(88)90013-5
- Weinitschke, S., Denger, K., Cook, A. M., and Smits, T. H. (2007). The DUF81 protein *TauE* in *Cupriavidus necator* H16, a sulfite exporter in the metabolism of C2 sulfonates. *Microbiol* 153, 3055–3060. doi: 10.1099/mic.0.2007/009845-0
- Winnen, B., Hvorup, R. N., and Saier, M. H. (2003). The tripartite tricarboxylate transporter (TTT) family. *Res. Microbiol.* 154, 457–465. doi: 10.1016/s0923-2508(03)00126-8
- Xu, Z., Nie, P., Sun, B., and Chang, M. (2007). Molecular identification and expression analysis of tumor necrosis factor receptor-associated factor 2 in grass carp *Ctenopharyngodon idella*. *Acta Biochim. Biophys. Sin.* 39, 857–868. doi: 10.1111/j.1745-7270.2007.00355.x
- Yellaboina, S., Goyal, K., and Mande, S. C. (2007). Inferring genome-wide functional linkages in *E. coli* by combining improved genome context methods: comparison with high-throughput experimental data. *Genome Res.* 17, 527–535. doi: 10.1101/gr.5900607
- Yu, N. Y., Wagner, J. R., Laird, M. R., Melli, G., Rey, S., Lo, R., et al. (2010). PSORTb 3.0: improved protein subcellular localization prediction with refined localization subcategories and predictive capabilities for all prokaryotes. *Bioinformatics* 26, 1608–1615. doi: 10.1093/bioinformatics/btq249
- Zarebinski, T., Hung, L. W., Mueller-Dieckmann, H. J., Kim, K. K., Yokota, H., Kim, R., et al. (1998). Structure-based assignment of the biochemical function of a hypothetical protein: a test case of structural genomics. *Proc. Natl. Acad. Sci. U.S.A.* 95, 15189–15193. doi: 10.1073/pnas.95.26.15189
- Zhang, T., Tan, P., Wang, L., Jin, N., Li, Y., Zhang, L., et al. (2017). RNAlocate: a resource for RNA subcellular localizations. *Nucleic Acids Res.* 45, D135–D138.
- Zmijewski, M. A., Kwiatkowska, J. M., and Lipińska, B. (2004). Complementation studies of the *DnaK-DnaJ-GrpE* chaperone machineries from *Vibrio harveyi* and *Escherichia coli*, both *in vivo* and *in vitro*. *Arch. Microbiol.* 182, 436–449. doi: 10.1007/s00203-004-0727-8

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Bharathi, Senthil Kumar and Chellapandi. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.